Homework 5

IE 7275 Data Mining in Engineering

Before you start: Read textbook KNN and Naive Bayes.

Submission Requirement: You should submit two answer sheets for this homework. One for non-coding problems 1 and 3 and the other for coding problems 2 and 4. Please type your steps and answers for the non-coding problems. Hand-written solutions will not be accepted.

Problem 1

The file Problem 1.xlsx contains 20 records of different features of pets. The response variable is *Pet class*.

a. Build a joint conditional probability table for all possible combinations of predictor and response value. For example, the table should present $P(Fur\ color = B, Fur\ type = F, Sharpness\ of\ claws = L \mid Pet\ class = ?)$ and so on.

b. Build a class conditional probability table of each predictor variable. For example, the table should present $P(Fur\ color\ = ? \mid Pet\ class = ?)$ and so on.

c. Consider a new record Fur color = W, Fur Type = F, and Sharpness of claws = L. Use both Exact Bayes and Naïve Bayes to classify this record with a cutoff value of 0.5.

Problem 2

Please refer to Google Colab file Homework 5 - Coding Problems. KNN

Problem 3

Please refer to Google Colab file Homework 5 - Coding Problems. - KNN

Problem 4

Please refer to Google Colab file Homework 5 - Coding Problems. – Naïve Bayes

Problem 5

Answer the following short answer questions and back up your answer with explanations and/or examples.

TODO 1

- What type of input and response variables can a Naive Bayes classifier handle?
- What kind of dataset is ideal for applying the Naive Bayes classifier?

TODO 2

- What are the pros and cons of a Naive Bayes classifier compared to other classifiers we learned in class?